Nonlinear and complex systems

- A nonlinear system is made up of a number of interacting components, which jointly determine the state (outcome) of the system.
- Nonlinear systems are the norm in biology.
- Nonlinear systems can be *complicated* (meaning they have a lot of components) and also *complex* (meaning they exhibit surprising or difficult to fathom behaviors).
- *Emergent behavior*: so complex as to be counter-intuitive.

E.g., a system of independent agents following two simple rules: (1) stay close to two nearest neighbors; (2) move away from all others; has the emergent property of forming a cycle.

Models

A model is anything that we use as a substitute (proxy) for a system we wish to understand.

The model embodies proposed key features and omits the rest.

- Physical model - ball-and-rod model of the DNA, animal models of human diseases
- Implicit model – learned through personal experience, refined through discussions

  Biological hypotheses are the *outcome* of such modeling

  Implicit models in molecular biology are often the mixture of qualitative biological considerations, mathematical description, and computational visualization/analysis.
- Declarative model – explicitly expressed as statements

  Declarative computational models use equations and algorithms, the tools of logic, statistics, etc.

The usefulness of computational models in molecular biology

The ultimate goal of gene regulatory network modeling is a mechanistic, fully predictive explanation of how information encoded in DNA underlies processes such as embryonic development and cellular form and function.

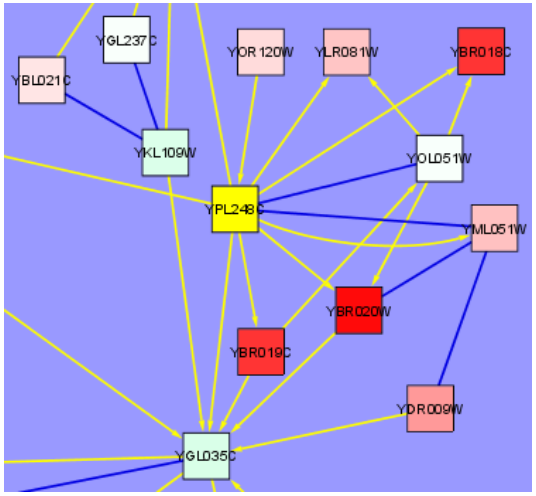
- A critical use of models in research is to predict how the system of interest will behave under novel conditions.
- Models may also be designed for communication of concepts, for sharing with collaborators, for re-use as components of larger models, and for training purposes.
- Tools developed by computational biologists can be developed into robust, easy-to-use software applications.
- Computational models permit the use of sophisticated analyses and visualization methods that reveal deeply hidden properties of systems.

Benefits and pitfalls of models of biological systems

- Mathematical and computational models can be unambiguously described and communicated.
- Making assumptions explicit can identify knowledge gaps and trigger new lines of investigation.
- In-silico experiments (i.e. thought-experiments facilitated by the power of computers) can explore scenarios too costly or too complicated to explore in the lab. They can help us develop insights into the roles of different elements and interactions, and guide experimental planning.
- Pitfalls: extrapolation may be risky; hidden assumptions, wishful thinking, situational bias, unknown unknowns, cascading errors, incorrect application
- Good practice: honesty, inclusiveness, openness

Implicit modeling via interaction network maps

Use of computational tools to visualize and analyze gene regulatory networks annotated with multiple types of data



protein-DNA int.

protein-protein int.

red: upregulated

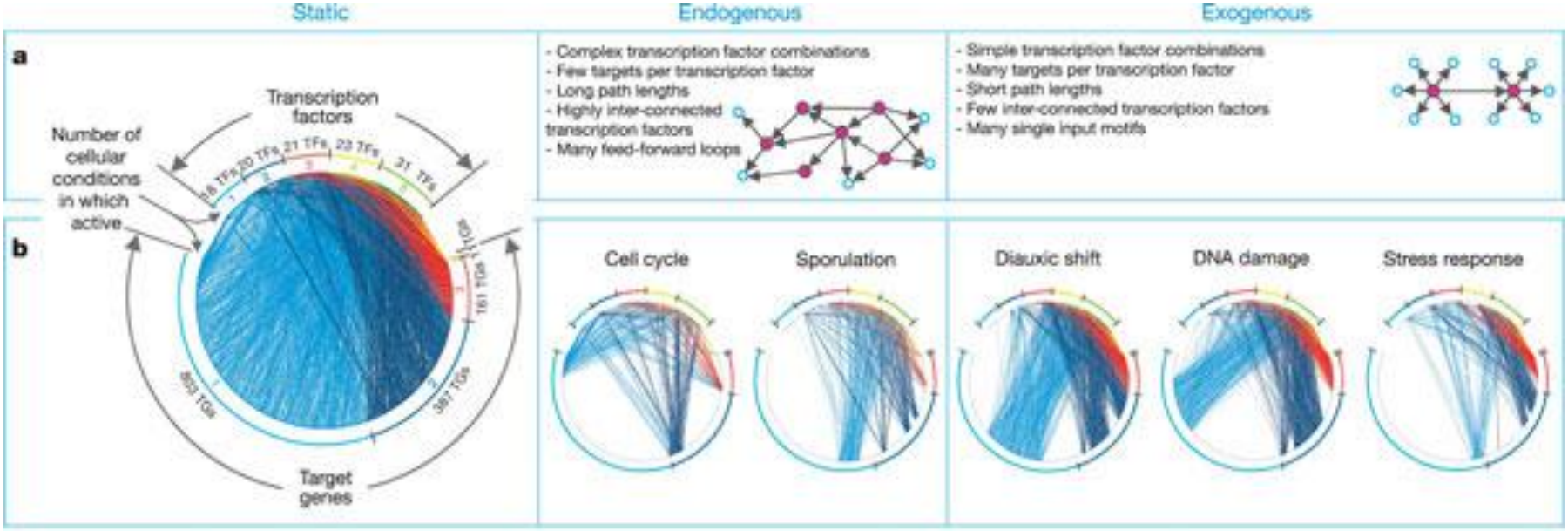
green: downregulated

Intensity shows fold change

yellow: selected

- Nodes can be clustered by shared gene ontology (GO) annotations or by community finding methods.
- The network characterizes the multiple roles of each gene.
- It helps predict the ways in which a perturbation can manifest itself.
- Gives insights into the global organization of cellular processes.

Condition-dependent transcription sub-networks

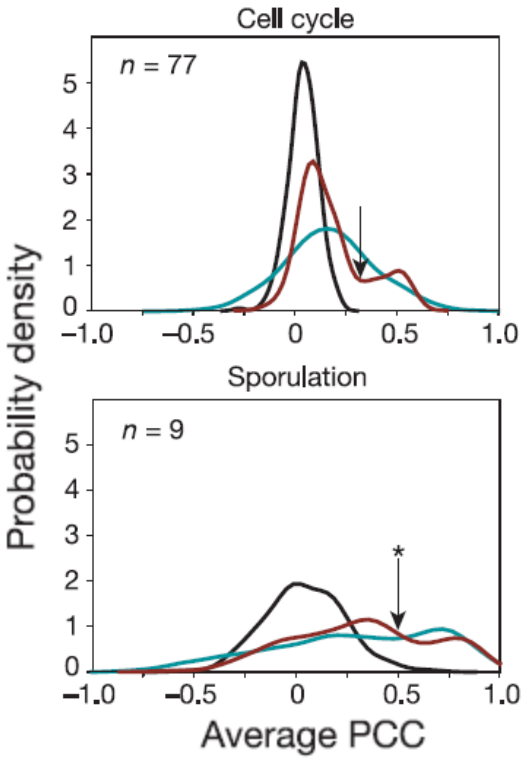


Endogenous
- Complex TF combination
- Few targets per TF
- Long path length
- Inter-connected TF
- Many FFL

Exogenous
- Simple TF combination
- Many targets per TF
- Short path length
- Few inter-connected TF
- Single input motifs

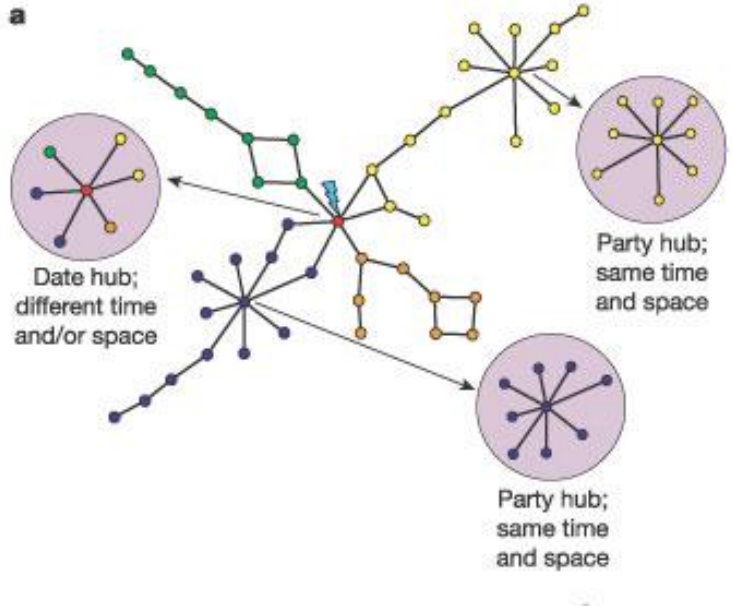
Luscombe et al, Nature 431, 308 (2004)

Not all protein-protein interactions are simultaneously active



Calculate the correlation between the expression time-course of genes encoding the first neighbors of hub proteins.

Two peaks – two different types of hubs.
Party hubs are inside connected modules that interact simultaneously. **Date hubs** connect different modules, their interactions happen at different times.

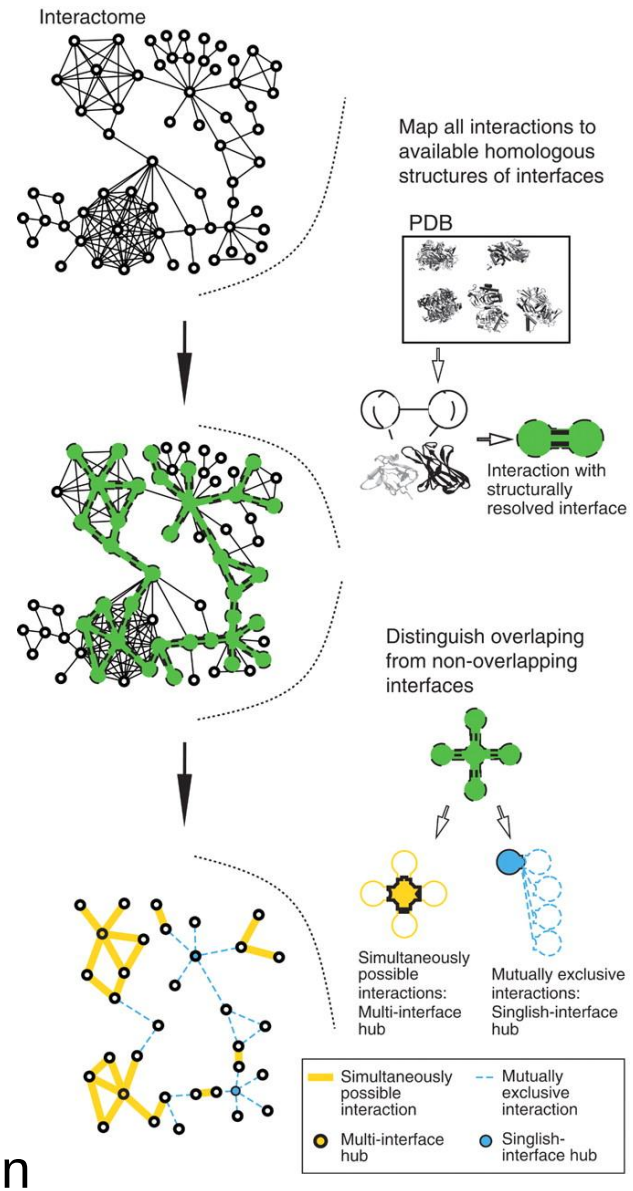


Han et al, Nature 443, 88 (2004)

Biological explanation of date/party hubs

Hub proteins are either multisite (multiple simultaneously possible interactions) or single site (mutually exclusive interactions).

Kim et al. Science 314, 1938 (2006)



Q: What relationship do you expect between date/party hubs and single/multi-site proteins?

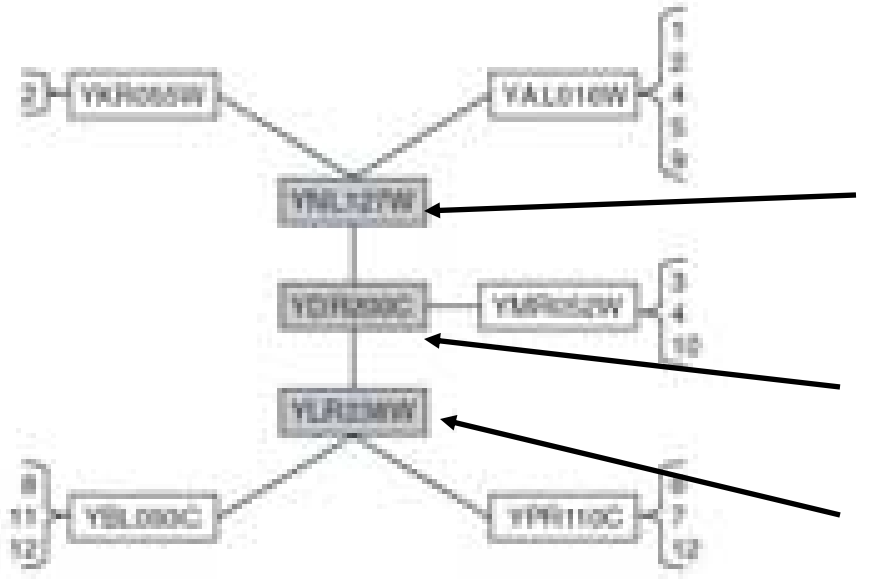
Interaction maps as predictive models

- Predict functional category – “guilt by association”
- Predict novel pathway members based on network clusters
- Predict novel interactions, or assign signs to regulatory interactions, based on functional or co-expression clusters
- Support or filter predictions made by other means

Functionality of proteins

- Assignment of functional category to proteins based on the functional category of their neighbors. Iterative process.

Subgraph of protein interaction network.



Functional classification

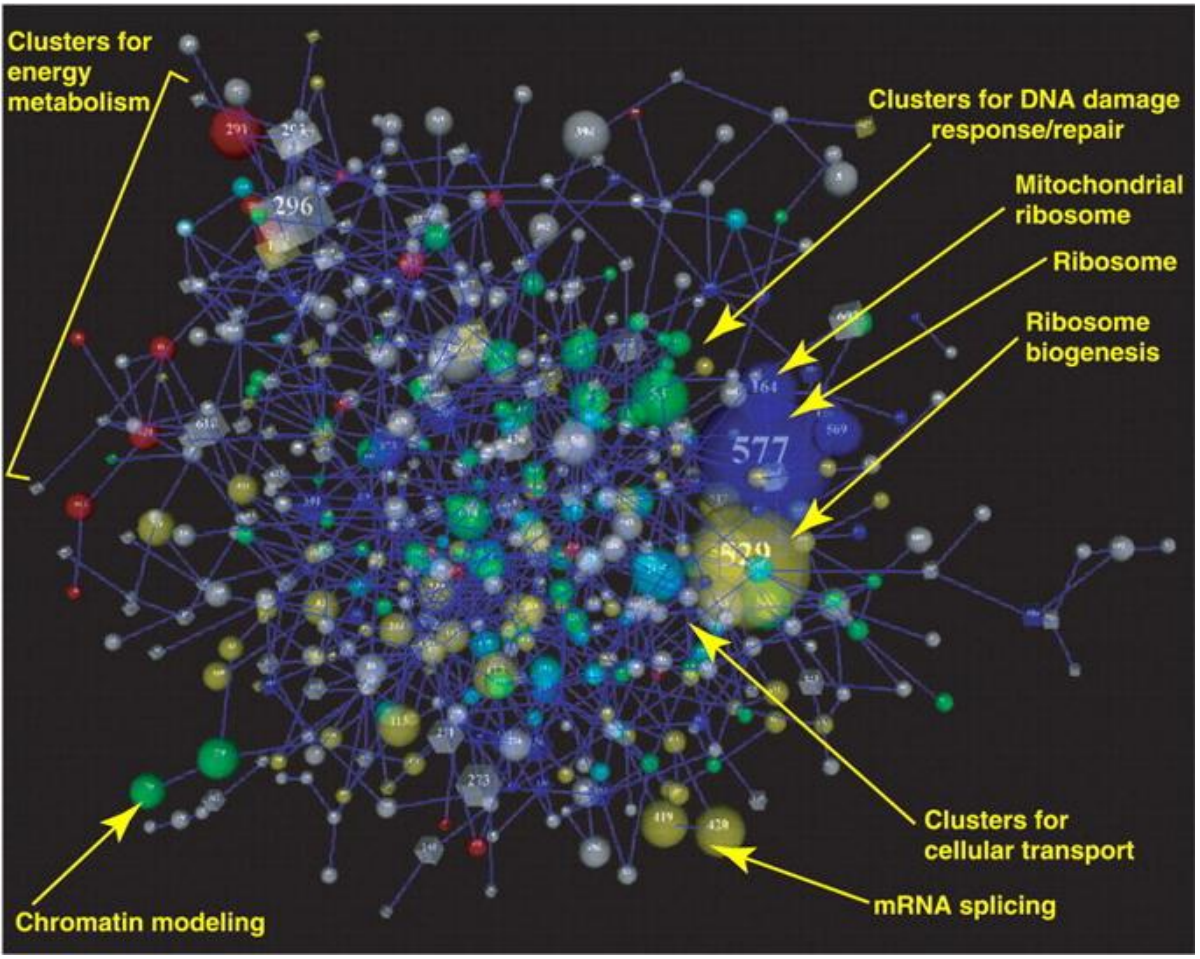
(2) – budding, cell polarity, filament formation

(3, 4, 10) – pheromone response, sex-specific proteins, cell cycle

(12) – nuclear organization

Lee, et. al. 2004 Science (306) 1555-58
Vazquez, et al. 2003 Nat. Biotech. (21) 697-700

Functional assessment of yeast genes

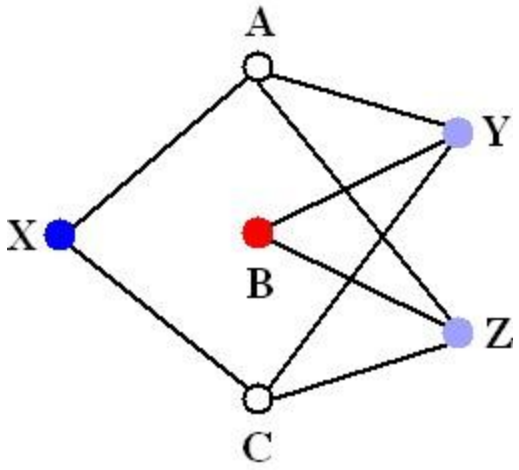


- Data used: Co-expression, gene fusion, phylogenetic profiles, co-citations, protein interactions.
- Likelihood of functional linkages estimated by their ability to reconstruct known gene pathways
- Found strongly connected functional modules

Inter-module linkage network

Nodes – modules, symbol size proportional to number of genes in the module.
Connections – inversely proportional to the fraction of genes linking the clusters.

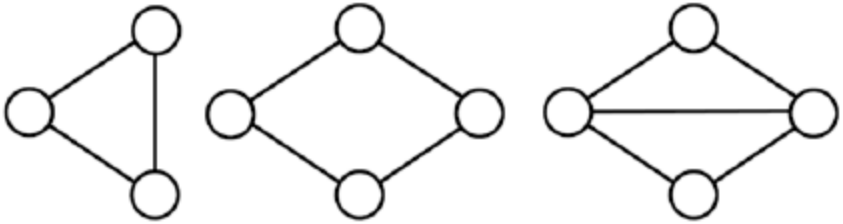
Prediction of protein interacting partners



- the interaction pattern of a protein forms a signature
- match it against all known signatures
- suggest as interaction partners the signature elements that the most similar proteins have, but the target protein does not.

Similar idea as in recommender systems

The method works due to the abundance of short cycles in protein interaction networks.



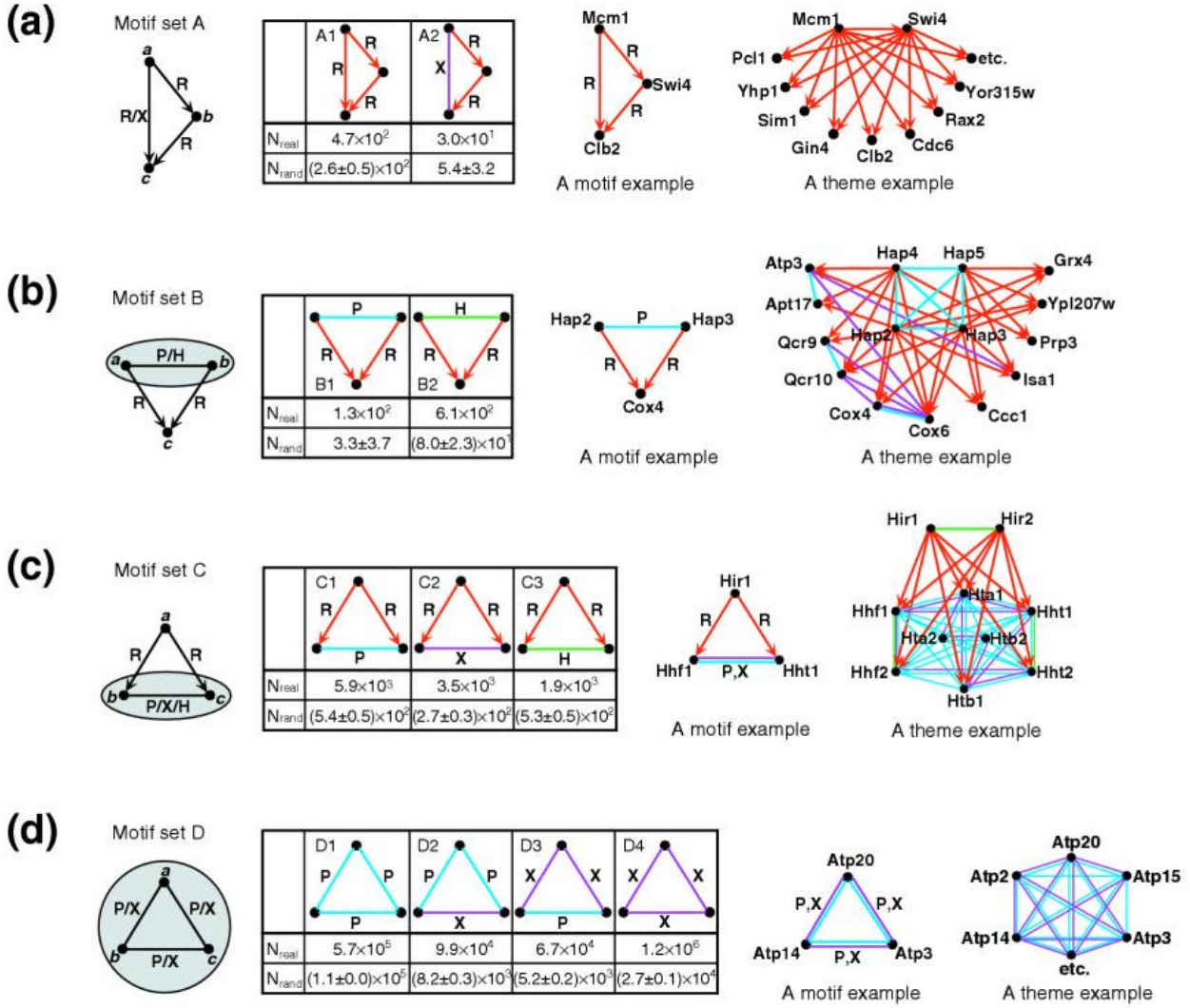
Albert et al. 2004 Bioinformatics (18) 3346-52

Interaction maps as predictive models (continued)

- Stereotypic interaction patterns (e.g. network motifs) provide functional insight
  - Positive feedback loop supports multistability
  - Negative feedback supports homeostasis
- Conservation of whole pathways predicts their importance
- Change of resolution by representing groups as metanodes allows switching between molecular detail and overall organizational principles.
- Incorporation of dynamic aspects as sequences of snapshots predicts cause-and-effect relationships.

Regulatory themes

R: Transc. reg
P: Prot. interaction
H: Seq. homology
X: Correlated expression



Zhang et al, J. Biol 4, 6 (2005)

General Topological Properties of Biomolecular Networks

A- power-law degree distribution

B- party hubs and date hubs

C- multi-site and single-site hubs

D- power-law distribution of branched pathways

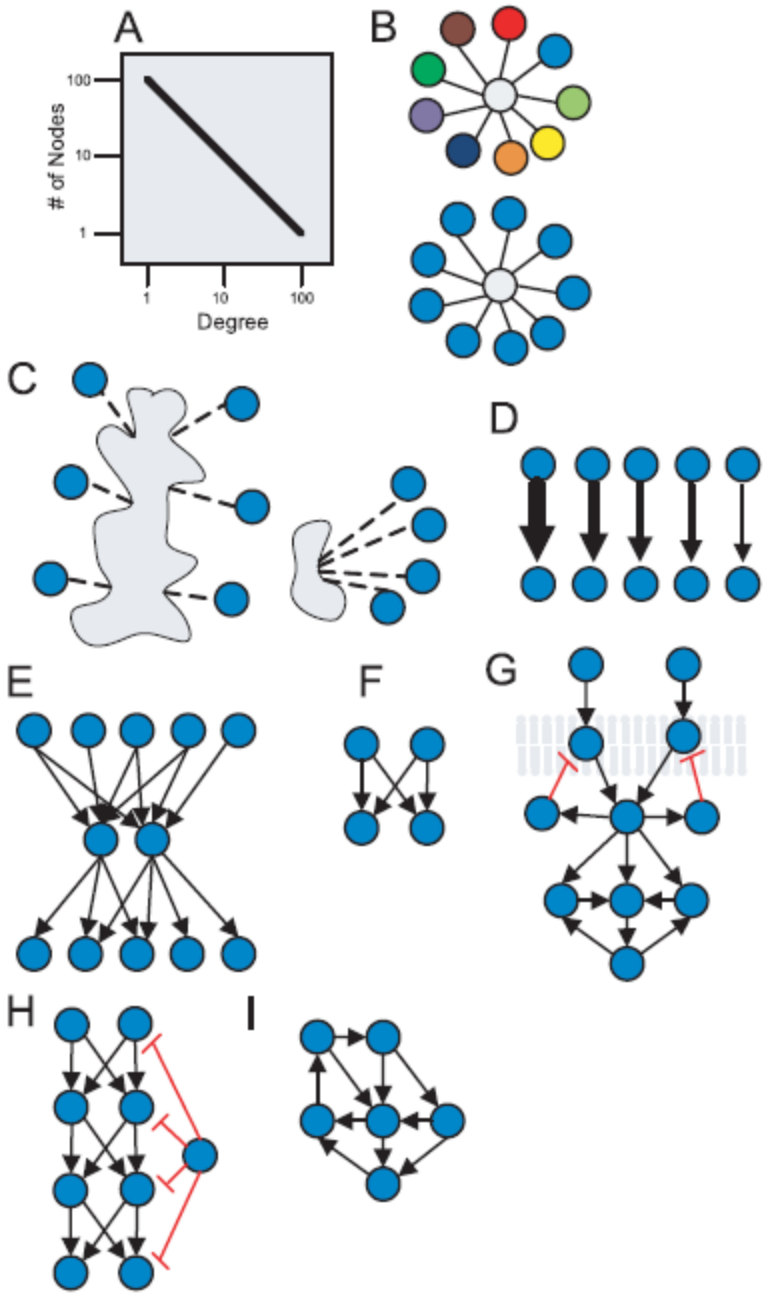
E- bow-tie structure of signaling pathways

F- bifans, the most common motifs

G- negative feedback loops at the membrane

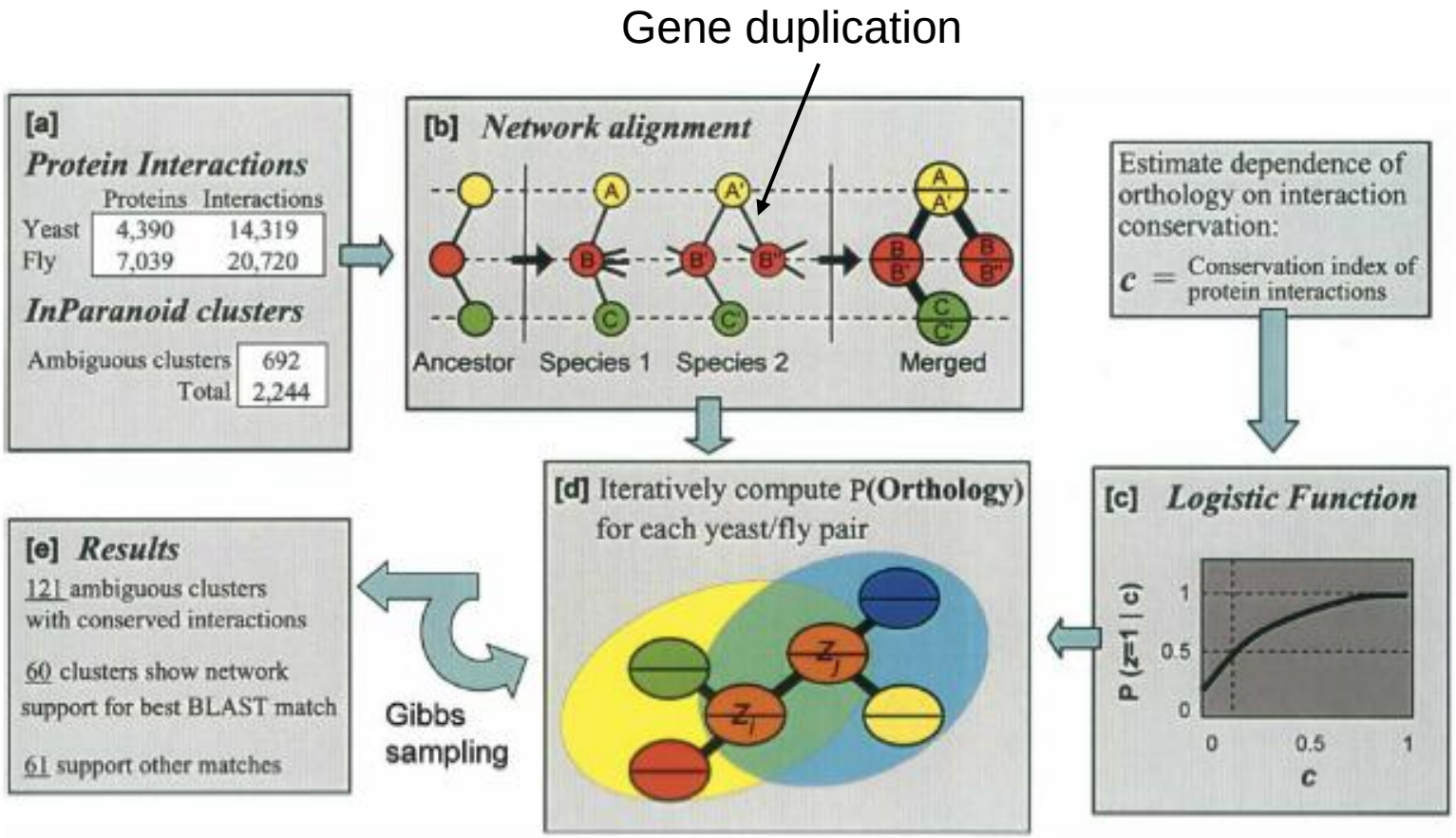
H- monotone system topology

I- nesting of positive feedback loops



Ma'ayan A. J Biol Chem. 2009 284(9):5451

Network alignment of yeast and fly protein-protein interaction networks to find functional orthologs



Only conserved interactions are used.

The network supports functional orthology for 61 pairs that are not the most conserved.

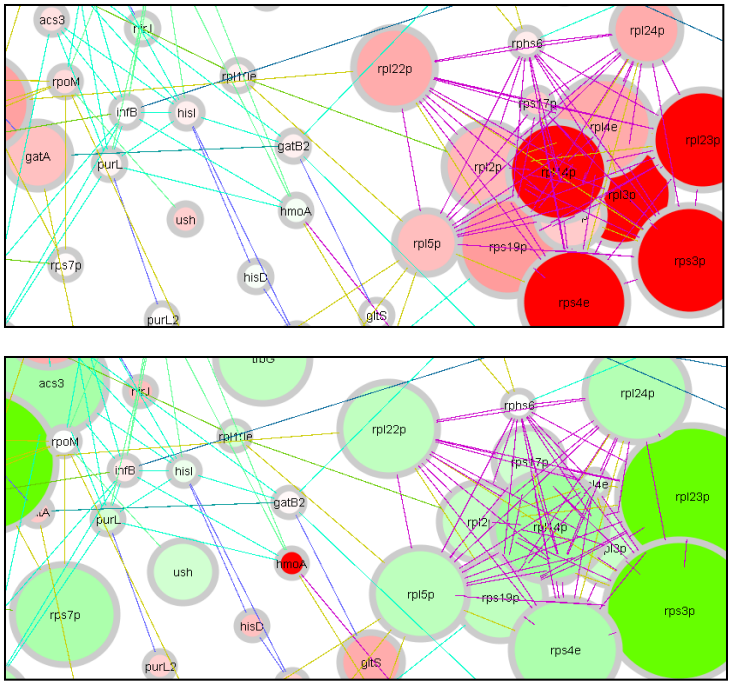
Bandyopadhyay, et. al. 2006 Genome Research 16: 428-35

Dynamic insights from visualization

Portion of Halobacterium NRC-1's interaction network, includes multiple interaction/relation types.

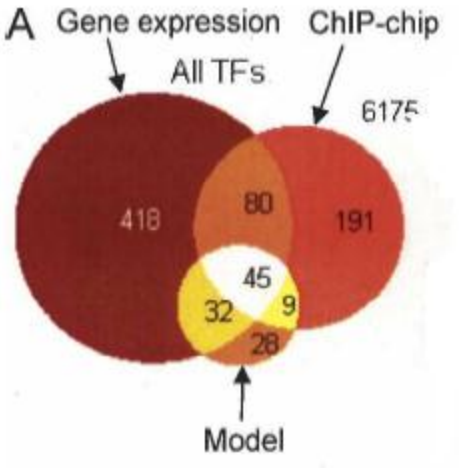
Two conditions shown.

Red indicates upregulation,
Green downregulation,
color intensity and node size indicate fold change.

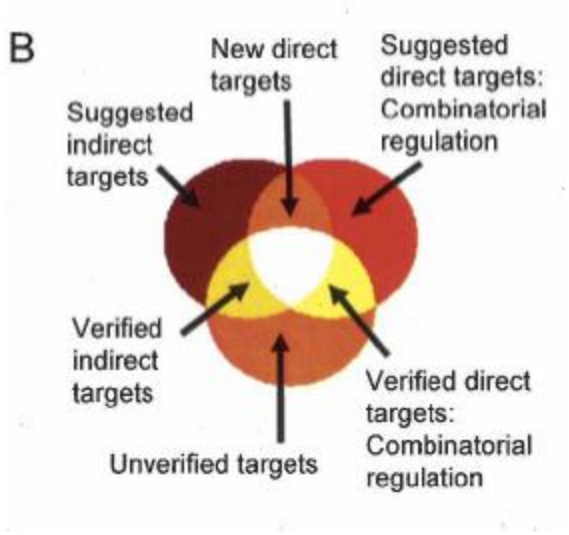


Identifies a subnetwork of highly co-regulated genes who are regulated in an opposite manner in the two conditions.

Integrated analysis of regulatory and metabolic networks



Numbers indicate number of genes (/750) in each category.



- Reconstruction of transcriptional regulatory network (55 TF) and metabolic network (750 genes).
- Network expansion – regulatory cascade connecting a transcription factor to its targets using chip-chip and binding site motif data.
- Dynamic model: Boolean regulation and flux balance analysis
- Predicts growth phenotypes of TF KO
- The network suggested previously uncharacterized interactions between TFs.

Herrgard, et. al. 2006 Genome Res. 16: 627-635

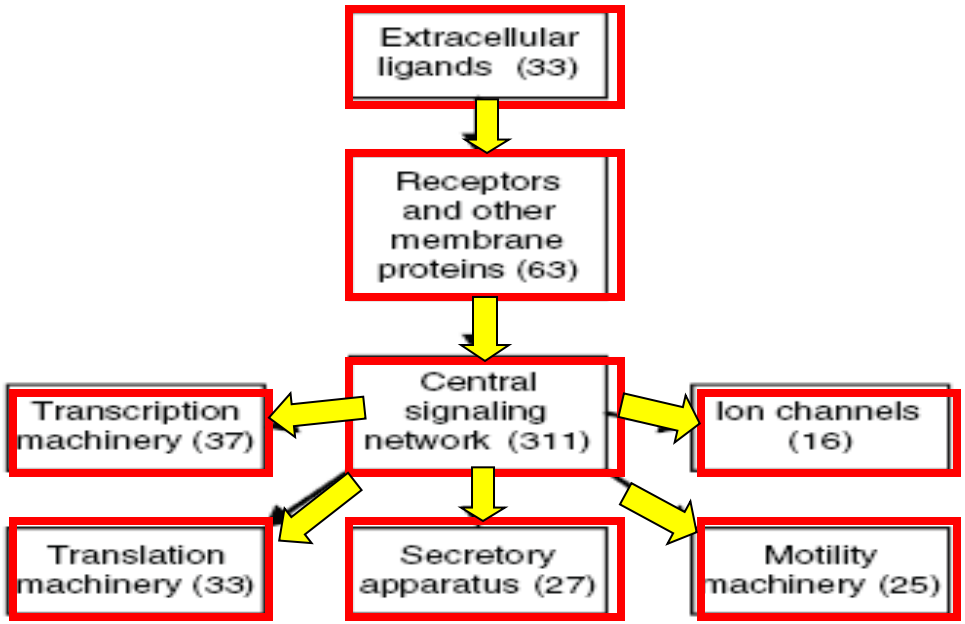
Signal transduction network of the hippocampal CA1 neuron

Data (binary interactions) collected form the experimental literature
System of interacting cellular components involved in phenotypic behavior

Edges can be directed or undirected (neutral)
Directed edges are activating or inhibitory

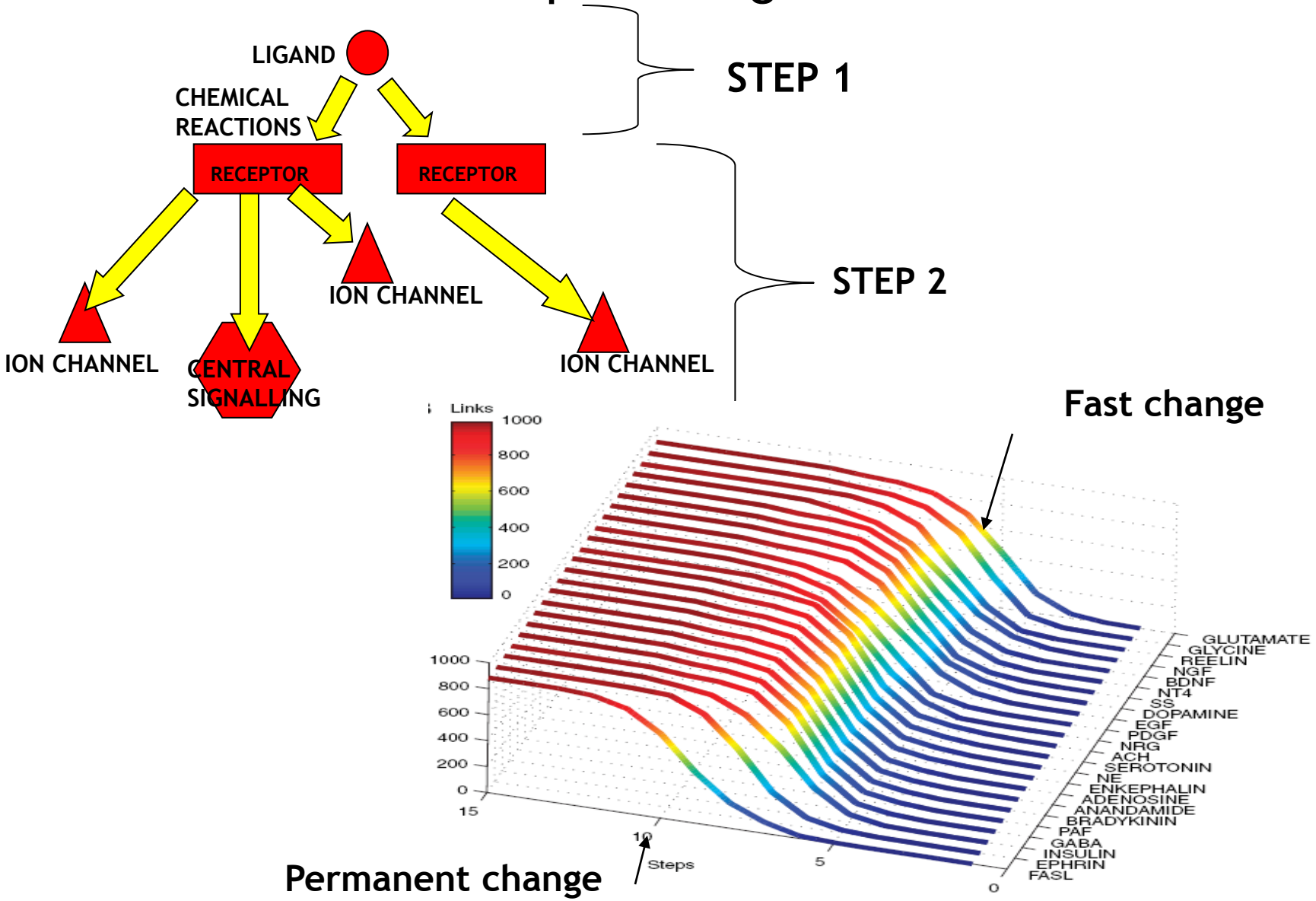
In and out degree distribution broad tailed

Highest degree nodes:
MAPK, CaMKII, PKA and PKC



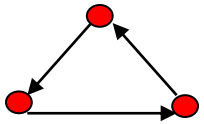
Ma'ayan et al, Science 309, 1078 (2005)

Signal propagation as links per step starting at a specific ligand

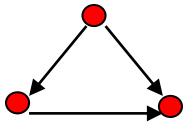


Motif abundance, homeostasis, and plasticity

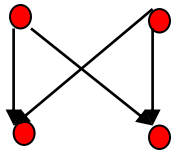
Feedback loops



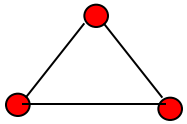
Feed-forward loops



bifans



scaffolds



Rapid-change ligands engage more motifs in fewer steps;

At early steps, more FFL than expected; at later steps, more +FBL

than expected

Positive and negative motifs are balanced for glutamate and BDNF - homeostasis;

More positive than negative FBL and FFL in NE – long- term info storage